MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Influence of recording technology on the determination of binaural psychoacoustic indicators in soundscape investigations

San Martín, Ricardo[a]; Arana, Miguel[b]; Ezcurra, Amaya[c]; Valencia, Ana[d];
Public University of Navarre, Science Department
Campus de Arrosadia s/n. 31006. Pamplona. Navarra. Spain

ABSTRACT

The evaluation of soundscapes is mainly carried out through field surveys, using soundwalking methodologies. Through rating scales and annotation of comments, the experiences and expectations of the participants are collected. Acoustic and psychoacoustic indicators are also reported to achieve a complete description of the acoustic environment. Binaural measurement systems should be used for registration in order to consider the way in which humans perceive the acoustic environment. Artificial heads or in-ear binaural microphones are the usual devices for this purpose. Further recording technology such as sound field microphones or microphone arrays are also used in soundscape investigations. These methods can provide a higher level of immersion in a later reproduction of the recorded soundscape through multi-channel setups. However, in contrast to direct binaural recordings, the determination of binaural psychoacoustic indicators requires to perform binaural synthesis by means of Head-Related Transfer Functions and signal processing. In this communication, the influence of different recording devices on obtaining psychoacoustic indicators such as binaural loudness and sharpness is analysed.

Keywords: Soundscapes, Psychoacoustics, Measurement techniques
I-INCE Classification of Subject Number: 72

1. INTRODUCTION

The term soundscape was adopted to give a holistic approach to the sound environment. The analysis of soundscapes aims to investigate all sounds perceived in an environment in all its complexity. This implies relating human perception with appropriate physical measurements. As specified in ISO 12913-2 [1], classical indicators are known to show strong limitations under certain sound conditions, as low frequency sound, tonal components or multisource environments. It may be necessary to choose different indicators depending on the soundscape under investigation. With respect to the representation of the sound environment, it is explicitly stated in the standard that it should be described using a combination of appropriate acoustic and psychoacoustic indicators.

[a] ricardo.sanmartin@unavarra.es,
[b] marana@unavarra.es,
[c] aezcurra@unavarra.es,
[d] ana.valencia@unavarra.es

The evaluation of soundscapes through field surveys is a costly process in time and depends on variables such as the concentration of the participants. The recreation of soundscapes in laboratory environments using virtual reality tools allows a more controlled evaluation, being able to focus the subject's attention on the acoustic aspects. New technologies offer tools to investigate specific variables related to perception in highly controlled environments, creating potentially multisensory configurations. However, the so-called ecological validity [2] of the data collected under such conditions has often been questioned.

In any case, a credible recreation of a soundscape must imitate our perception of the three-dimensional sound scene. Therefore, specific recording devices are necessary, as well as analysing their effect on the psychoacoustic parameters related to that perception.

1.1 Recording techniques

"Acoustic measurements related to a soundscape shall consider the way human beings perceive the acoustic environment [1]". Binaural audio recordings made with an artificial head are usually considered to provide the highest degree of realism. Through these devices, the soundscapes can be recorded as if one were present in the original acoustic environment. In that way, recorded binaural signals contain embedded spatial cues that mimic human perception. When correctly reproduced, they create a powerful impression of immersion in the natural listening environment.

However, there exists great variation in the shape and size of people's head and pinnae. With dummy head recordings, localization cues superimposed on the recorded sound correspond to averaged anatomical features. Using individualized binaural recording devices can lead to an improved experience, at least for the person involved. Binaural microphones consisting of small capsules located at the entrance to the ear can provide a high level of customized spatial perception, also adding portability to the procedure.

One of the constraints of recording with dummy head or binaural microphones is that the scene is fixed. Later in the reproduction, the listener cannot rotate the head in order to interact with the acoustic environment. This inconvenience can in theory be solved using multichannel recording techniques, as ambisonics [3]. Although developed in the 70s, ambisonics has gained weight recently since YouTube, Oculus VR and Facebook adopted it as a standard for their 360-degree videos. This technique can provide an alternative to binaural recordings in the context of studies of soundscapes [4], if the semantic aspects of user experience are similar in the original soundscape and its reproduction [2].

However, the ambisonics technique presents some disadvantages. Compared with binaural techniques, the level calibration and equalization processes are more complex. If playing through headphones, it is necessary to monitor the movement of the head, which entails real-time HRTF processing. Also, the spatial resolution of first order ambisonics is low. And higher order ambisonics recording require sophisticated equipment. Non-linear techniques as directional audio coding (DirAC) [5] can increase the quality of reproduction when compared to signal-independent decoding of the same first-order ambisonics input signal.

Other techniques employ microphones in form of arrays. These techniques are widely used in location and separation of sources, noise reduction, echo cancellation, etc. The recordings of the microphones are not used directly but require further processing. For example, by processing the audio input from two microphones the predominant direction of arrival (DOA) can be estimated. Then, in a following stage,

the binaural synthesis is performed to create a signal that imitates the human hearing system response by using HRTFs and the DOA information resulting from the previous stage [6].

1.2 Psychoacoustic parameters

Psychoacoustic parameters represent a major role with respect to auditory sensations. These parameters allow obtaining information with greater differentiation than when considering the sound pressure only. Beyond the purely physical sound level in decibels, loudness parameter considers human signal processing effects like frequency weighting, frequency and temporal masking, critical bands and other nonlinearities related to the mechanism of the cochlea. It shows a higher correspondence with the auditory sensation of level ordered on a scale from quiet to loud.

The unit of loudness is the sone. It is defined by stating that a loudness of 1 sone is equivalent to the loudness of a 1 kHz tone at a sound pressure level of 40 dB. A sound that is twice as loud as another sound is characterized by doubling the number of sones. To calculate the loudness parameter, different procedures have been developed. ISO 532-1 allows for determining the loudness on both stationary and non-stationary signals [7,8]. Specific loudness N' exhibits the distribution of loudness across the critical bands. The total calculated loudness N is the result of the specified loudness values N' through integration of the critical band rate. It corresponds to the loudness that would be experienced by an average of a group of persons with ontologically normal hearing whose heads are centred at the position of the microphone.

The sensation of sharpness is a dimension of sound character related to its spectral content. It is related with the location of the centre of gravity of the amplitude spectrum, increasing its value when high-frequency components are present. The sharpness parameter S is calculated adding a weighting function *g(z)* to the specific loudness spectrum [9].

$$S = 0.11 \frac{\int_0^{24} N'(z) g(z) z dz}{\int_0^{24} N'(z) dz} \text{ acum} \qquad \text{(equation 1)}$$

The unit of sharpness, the acum, is defined such that narrowband noise in the critical band of 1 kHz at a sound pressure level of 60 dB. Procedures for computing the sharpness are standardized in DIN 45692 [10]. There is an alternative calculation method that differs slightly in the weighting function [11]. Also, as psychoacoustic tests reveal that the level of sound moderately affects the perceived sharpness, other methods apply a certain total loudness dependent weighting [12].

A model for combining dichotic situations causing interaural differences into one global sensation is missing. When presented to both ears, a sound is perceived as higher than monaurally presented. Some studies suggest an increase of 1.5 in the perceived loudness [13] and propose a combined calculation for the binaural situation based on the loudness of each channel. With respect to binaural sharpness, the same proportion value as in binaural loudness situation is usually adopted [14].

2. PRESENTATION OF SOUNDSCAPES IN LABORATORY

To carry out the experiences described in the following section, the maximum control over the stimuli presented to the different recording devices was prioritized [15]. Two types of soundscapes were used: synthesized and recorded. Figure 1 graphically

shows the process. The stimuli consist either of an original environment where the signals were recorded with a sound field microphone or a virtual scene in which spatialization was parameterized. The soundscapes obtained are transferred to the destination environment where they are rendered using loudspeakers. In the last stage, the soundscapes are evoked in the brain of the subject at the same time that their characteristics can be evaluated with laboratory equipment.

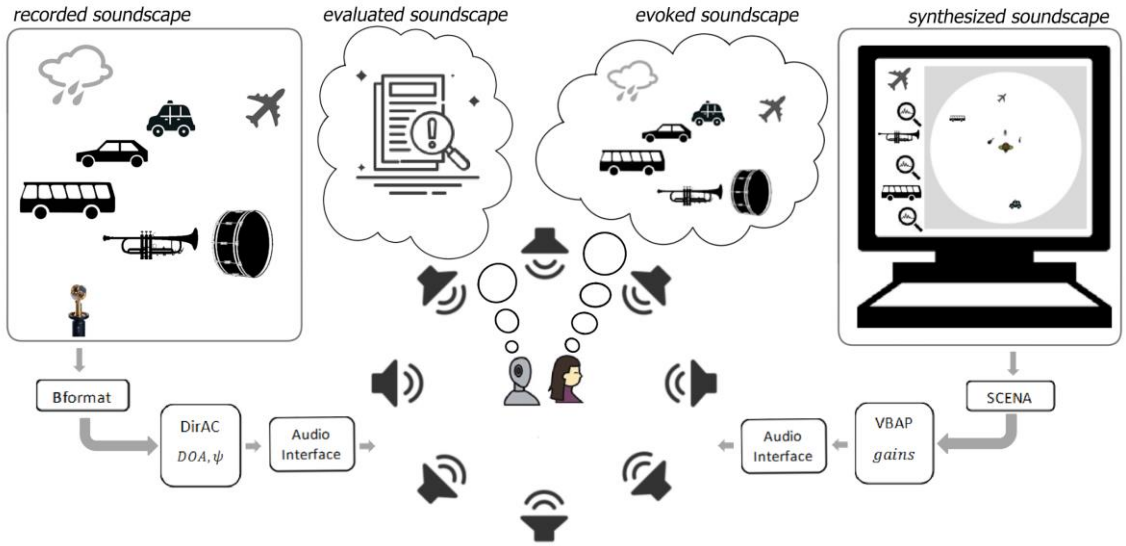

Figure 1. Presentation procedure for recorded and synthesized soundscapes.

2.1 Synthesized soundscapes

A tool called SCENA has been developed that allows to recreate sound scenes in laboratory environments by combining spatial audio processing with the possibility of recording the result that is generated in real time [16]. It has a user interface that, on the one hand, allows to configure new scenes and loudspeakers setups in a simple way and, on the other hand, shows the virtual scene with basic player functions as play, pause, stop and rec.

The software has been designed in order to create virtual sound scenes from the loading of audio files. Each of the loaded tracks is represented in the virtual scene by means of an icon that can be moved in real time, in addition to changing its status to active or inactive. Also, SCENA gives the option to export the resulting sound scenes in different formats, in order to use them for further research.

The total development of SCENA has been carried out using Matlab software, from the user interface through which the virtual scene is configured and controlled up to the communication with the audio reproduction hardware. SCENA allows different spatial processing algorithms. This time vector-based amplitude panning (VBAP) was used on the multichannel setup.

2.2 Recorded soundscapes

As a second option, soundscapes were recorded by means of a first order ambisonics microphone - Sennheiser Ambeo - equipped with four cardioid capsules positioned in the form of a tetrahedron. The signals obtained (4ch A-format) were stored in a multitrack recorder - F4 zoom - previously calibrated in the laboratory.

The conversion to B-format signals was done by means of a VST plugin provided by the manufacturer, finally obtaining a signal (W) proportional to the pressure (omnidirectional) and three to the pressure gradient in each XYZ axis (figure

of 8). In this way, the particle velocity vector at the recording point can also be estimated.

From the information relative to pressure and velocity, the intensity vector of the sound field was obtained, which expresses the direction and magnitude of the sound energy flow. By analyzing this vector, the predominant direction of arrival (DOA) and diffusion parameter ψ can be estimated. From these parameters and from the WXYZ signals coming from the B-format microphone, the synthesis of the signals sent to the speakers was performed [17].

3. EXPERIMENTAL PROCEDURE

Audio recordings were performed at the listening room of the Acoustics Laboratory of the Public University of Navarre. It has a reverberation time of about 0.2 seconds at mid-frequencies. Synthesized and recorded soundscapes were emitted through a Focusrite Scarlett 18i20 audio interface and eight Neumann KH120A loudspeakers. The capture of audio signals with the evaluated devices was made at the center point of a circular speaker configuration, at a height of 1.7m. Four different soundscapes were evaluated. Two of them recorded (S1 and S2) and the other two synthesized (S3 and S4).

3.1 Devices

The different setups consist of the following components. For direct binaural audio recordings two different systems were used: HEAD acoustics HSU III artificial head together with Norsonic type 335 preamplifier and signal conditioner, and Roland CS-10EM binaural microphones along with Zoom H2n handy recorder.

For first-order ambisonics, Sennheiser Ambeo VR mic composed by four cardioid capsules in tetrahedral arrangement and Zoom F4 multitrack field recorder.

A low-cost acquisition system was also evaluated. It consisted of an array of four microphones belonging to a Sony PlayStation Eye Camera, with a sampling rate of 16 kHz and a distance between the outermost microphones of 62mm.

Finally, for calibration purposes, an omnidirectional GRAS microphone was also included among the evaluated recording devices.

3.2 Calibration

The parameters that are to be measured, such as loudness, depend on the sound level, which is why a correct calibration of the recording system is critical. This task is relatively simple with an acoustic calibrator that uses a closed coupling volume to generate a precise sound pressure on an omnidirectional microphone.

However, the calibration of binaural recording systems is a more complicated issue. By definition, a binaural recording should modify a sound field like a human, according to the human anatomy. These changes can or cannot be direction-dependent. Therefore, the sound level obtained will depend on the spectral and directional characteristics of the sound field used for the calibration.

On the one hand, the modifications induced to the sound waves from the cavum concha to the eardrum are independent of the direction of incidence of the sound waves. When listening a binaural recording obtained with an artificial head it is advisable to use equalization to correct the fact that the headphones cannot be placed in front of the eardrum. Otherwise, the sound waves would travel the auditory canal “again”.

On the other hand, the head, shoulders and outer ear influence sound fields in a direction-dependent way. In order to make binaural recordings and conventional

omnidirectional recordings compatible in a way they can be compared, equalization is also needed. However, in this case, equalization characteristics should depend on the type of sound field recorded, since different direction dependent modifications can happen.

Keeping all this in mind, a calibration procedure was designed. First, uncorrelated white noise (S0) was emitted by the 8 speakers and the sound field recorded with the calibrated omnidirectional microphone. Then, a "diffuse-field" equalization was applied to the signal to take into account the effect of human anatomy under those conditions. This equalization considers the direction dependent factors averaged over all directions of sound incidence as well as the direction independent factors. The level L_{Aeq} thus obtained was established as a reference. Finally, the same noise was registered with the rest of the systems and signal processing for binaural rendering, if needed, applied. For each system, the average of the levels obtained in both ears was matched to the reference level by a calibration factor. Later this factor was applied to all soundscapes analyzed (S1-S4). Table I shows the L_{Aeq} registered for all devices evaluated. Differences up to 6 dB can be found.

Table I. A-weighted equivalent continuous sound level L_{Aeq} (dB) of the different soundscapes and recording devices evaluated

	S0	S1			S2			S3			S4		
	Av	L	R	Av	L	R	Av	L	R	Av	L	R	Av
AH	62.1	41.7	48.2	44.9	47.3	44.9	46.1	48.6	50.4	49.5	45.9	51.2	48.5
AX	62.1	46.0	52.4	49.2	49.8	48.0	48.9	50.2	51.7	51.2	49.1	54.1	51.6
BA	62.1	46.6	52.5	49.6	48.1	47.8	48.0	49.4	51.8	50.6	50.0	55.7	52.8
BD	62.1	43.0	49.3	46.1	47.5	46.2	46.9	46.0	51.4	48.7	47.7	54.3	51.0
BM	62.1	46.9	52.5	49.7	50.2	48.3	49.3	50.1	51.3	50.7	50.0	53.1	51.6
OM	62.1	47.7	47.7	47.7	47.6	47.6	47.6	47.4	47.4	47.4	49.2	49.2	49.2
PS	62.1	45.1	45.1	45.1	45.8	45.8	45.8	51.1	51.1	51.1	50.7	50.7	50.7

3.3 Binaural rendering

Soundscapes registered with the artificial head (AH) were analysed with the only modification relative to their calibration factor, and the same consideration regarding binaural microphone. With respect to this last device, two records were taken: one with the device located in the ears of the artificial head (BM) and the other with the device on a subject (AX).

Some of the recording devices used do not present a binaural signal as an output. Consequently, binaural rendering was performed from the output signals of each device. Different strategies were adopted. Mono recordings (OM) were filtered with the horizontal diffuse-field response of the artificial head. This filter was formed as the power average across a set of 72 measurements taken around a uniform circular distribution of horizontal directions.

With respect to the low-cost array (PS), one of the channels was chosen and the same procedure than for mono recordings was applied. Strictly speaking, if you want to use this device to capture spatial audio for greater immersion in presentation, you need to process the signal captured by at least two of their microphones. By exploiting the estimated directional characteristics of the incoming sound by means of DOA estimation, the binaural rendering is conducted lately through HRTF filtering.

Finally, the sound field microphone signals were processed considering that the binaural rendering of any loudspeaker system can be achieved by convolving the HRTFs of the position of the loudspeakers with the audio fed to those loudspeakers. So, the approach consisted in creating a setup of virtual loudspeakers - matching the real setup of the listening room - and filter the virtual output channels with the corresponding static HRTF filters. Those virtual loudspeaker channels were rendered following two different approaches: ambisonics linear decoding (BA) and non-linear directional audio coding (BD).

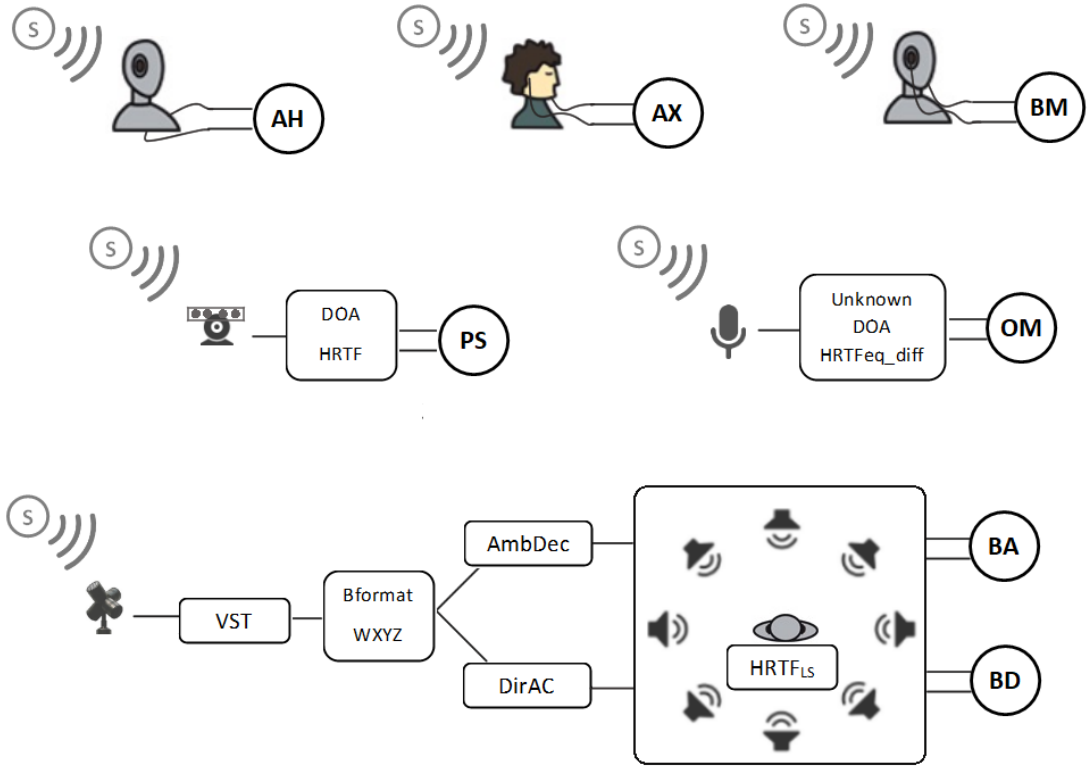

Figure 2. Binaural rendering corresponding to each recording device

In the first case, the virtual loudspeaker signals are derived by using a linear combination of the ambisonic component signals. This decoding matrix was generated by pointing a virtual supercardioid microphone in the direction of each virtual loudspeaker. In practice, a real ambisonic decoder requires some optimisations to work properly, but this approach was considered representative of a rough implementation.

In the second case, first-order directional audio coding (DirAC) was chosen as the representative of a new family of approaches that are explicitly based on psychoacoustic mechanisms. DirAC is a non-linear method in the time-frequency domain to reproduce spatial sound that exploits the fact that the mechanisms for the perception of diffuse sound are very different from those related to the perception of direct sound. In this way, a spatial audio signal is broken down into diffuse and non-diffuse components, and the two groups of signals are rendered using different

techniques. The non-diffuse (direct) part is reproduced as if it were a virtual point source located in the direction defined by the DOA value, applying VBAP, that is, by amplitude panning. The diffuse part is reproduced by all the speakers that surround the listener after applying decorrelation filters, which ideally would generate the so-called diffuse field. Figure 2 aims to graphically summarize all binaural renderings described.

4. RESULTS AND DISCUSSION

The soundscapes recorded by the different devices were analysed by means of BKconnect software. Stationary (one minute) sound quality metrics are shown in Figure 3. Values between 6.9 and 13.1 sones were obtained for binaural loudness and 1.81 and 2.95 acum for binaural sharpness.

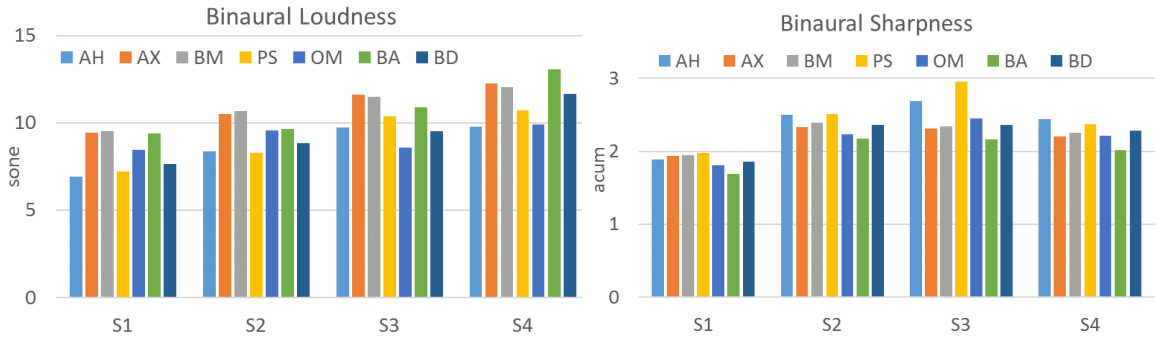

Figure 3. Values of the binaural psychoacoustic parameters loudness (left) and sharpness (right)

As expected, no significant differences were found between the AX and BM "devices". While the use of individualized HRTFs can increase the spatial impression in a later reproduction of binaural recordings, it does not seem to affect the obtaining of the psychoacoustic parameters.

The differences found between the BA and BD modes are more striking. The device used to record the sound field is the same in both systems. Thus, the differences are attributable to the posterior binaural rendering. Ambisonics linear decoding systematically get higher loudness and less sharpness. The differences can reach 1.7 sones and 0.26 acum respectively.

Systematic differences of up to 2.6 sone have also been found in the loudness parameter between direct recordings made with artificial head (AH) and binaural microphones (AX or BM). These differences could be expected from the values of L_{Aeq} (see Table I) since these were always greater for the binaural microphone. Another calibration process, as a sound field dependent one, could have decreased these variations.

Finally, it is worth noting the results obtained by the low-cost device (PS). If they are compared with the artificial head, the maximum differences obtained are 0.9 sone for stationary binaural loudness and 0.26 acum for sharpness. In this case, simple binaural rendering with calibrated HRTF diffuse equalization seems enough to reflect auditory sensations as loudness and sharpness. However, this issue needs further investigation since the results obtained with the omnidirectional microphone (OM), which follows the same procedure, show greater deviations from AH values.

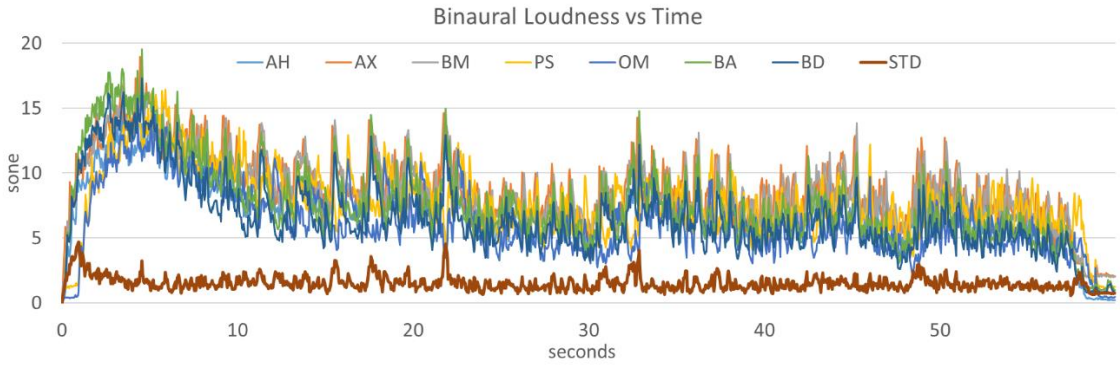

Figure 4. Binaural Loudness vs time, soundscape S3

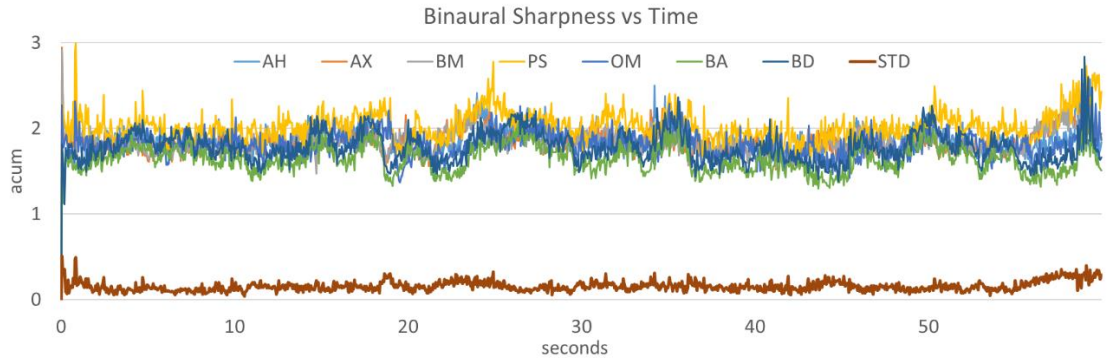

Figure 5. Binaural Sharpness vs time, soundscape S1

Figures 4 and 5 show the time dependence of the parameters. Values for loudness (Figure 4) and sharpness (Figure 5) obtained every 50 ms are plotted. Although the graphs for two specific soundscapes are shown (S1 and S3), the results were similar in the rest of the cases. The variability, evaluated as standard deviation, remains practically constant in both parameters although the value of the parameter varies, being around 1.5 sones for loudness and 0.15 acum for sharpness. Although more research would be advisable to analyse other possible variables, such as the orientation of the recording device, these values could be indicators of the uncertainty associated with the recording system used.

5. CONCLUSIONS

A procedure has been designed to analyse the influence of different recording devices in obtaining psychoacoustic indicators. Significant differences have been found between recordings made with artificial head and binaural microphones. The calibration process has been revealed as a determining factor, at least for obtaining loudness.

The performance of capture devices other than binaurals has also been evaluated. These devices would allow a greater spatial impression in a later reproduction of the recorded soundscape. However, the signal processing necessary to be able to compare it with a binaural system can introduce differences in the measured psychoacoustic parameters.

With proper equalization, low cost microphones can be used to evaluate the psychoacoustic parameters. Differences of less than 1 sone were obtained for loudness when compared with the results of an artificial head.

Finally, a first estimate of the uncertainty associated with the recording system has been determined. This would be around 1.5 sones for loudness and 0.15 acum for sharpness.

ACKNOWLEDGEMENTS
This work has been funded by the Spanish Ministry of Economy and Competitiveness through the R+D+I research project with reference BIA2016-76957-C3-2-R.

REFERENCES
1. ISO 12913-2 Soundscape: data collection and reporting requirements, ISO Technical Specification, Geneva, Switzerland (2018)
2. Guastavino C., Katz B.F.G., Polack J.D., Levitin D.J., Dubois D., Ecological validity of soundscape reproduction, *Acta Acustica United with Acustica 91(2)* 333-341 (2005)
3. Sun K., Botteldooren D., De Coensel B., Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation. Proceedings of the 47th International Congress and Exposition on Noise Control Engineering, Chicago, USA (2018)
4. Davies W.J., Bruce N.S., Murphy J.E., Soundscape Reproduction and Synthesis. *Acta Acustica United with Acustica* 100(2), 285-292 (2014)
5. Pulkki V. Spatial sound reproduction with Directional Audio Coding. *Journal of the Audio Engineering Society* 55(6), 503-516 (2007)
6. Cobos M, Lopez J.J., Spors S.,A Sparsity-Based Approach to 3D Binaural Sound Synthesis Using Time-Frequency Array Processing. *Journal on Advances in Signal Processing* 2010(1) 10.1155/2010/415840 (2010)
7. ISO 532-1: 2017, Methods for calculating loudness: Part 1 - Zwicker method. ISO Technical Specification, Geneva, Switzerland (2017)
8. ISO 532-1: 2017, Methods for calculating loudness: Part 2 - Moore-Glasberg method. ISO Technical Specification, Geneva, Switzerland (2017)
9. Fast H, Zwicker E. Psychoacoustics - Facts and models. Springer (2007)
10. DIN 45692:2009, Measurement technique for the simulation of the auditory sensation of sharpness. Deutsches Institut für Normung (2009)
11. Bon Bismarck G. Sharpness as an attribute of the timbre of steady sounds. *Acustica* 30: 159-172 (1974)
12. Aures W. Berechnungsverfahren für den Wohlklang beliebiger Schallsignale. Ein Beitrag zur gehörbezogenen Schallanalyse, Dissertation, TU München, Deutschland, (1984)
13. Moore B. C., Glasberg B. R., "Modeling binaural loudness", *Journal of the Acoustical Society of America* 121(3) 1604-1612 (2007)
14. Segura-Garcia J, Navarro-Ruiz JM, Perez-Solano JJ, Montoya-Belmonte J, Felici-Castell S, Cobos M, Torres-Aranda AM. Spatio-Temporal Analysis of Urban Acoustic Environments with Binaural Psycho-Acoustical Considerations for IoT-Based Applications. Sensors 18(3) 10.3390/s18030690 (2018)
15. San Martín R., Valencia A., Vicuña M., Ezcurra A., Arana M., Grabación y presentación de paisajes sonoros en entornos de laboratorio. Proceedings of the XI Congreso Iberoamericano de Acústica FIA 2018, Cádiz, Spain (2018)
16. Valencia A., San Martín R., Ezcurra A., Arana M., Aplicación en Matlab para el diseño y presentación de paisajes sonoros virtuales Proceedings of the XI Congreso Iberoamericano de Acústica FIA 2018, Cádiz, Spain (2018)
17. Vilkamo J, Lokki T, Pulkki V. Directional audio coding: Virtual microphone-based synthesis and subjective evaluation. *Journal of the Audio Engineering Society* 57(9): 709-724 (2009)